OUTLINE

• Documentation of research data at the GESIS Data Archive
  • Metadata in DDI-C and DataCite format
  • DOI persistent identifiers for research data
• Usage of DOIs for data citations
• Machine actionable processing of metadata
  • IsMetadataFor and HasMetadata properties in the DataCite metadata schema
  • Example implementation
• Lessons learned and other possible solutions
DOCUMENTATION OF RESEARCH DATA AT THE GESIS DATA ARCHIVE

• Documentation for the research data from the social sciences at the GESIS Data Archive is being done on several levels:
  • The Data Catalogue DBK has the main level of information for finding appropriate studies
  • Further documentation includes methodological reports, questionnaire and dataset documentation, English and German translations
  • The DBK metadata is compatible with DDI-C and DDI-L and can be exported
  • The DBK metadata is (mostly) compatible with DataCite and is transferred to da|ra when DOIs are registered
  • Further documentation is only available for parts of the collection which are documented in depth (with a Codebook). This documentation is produced internally in DDI-C compatible form, but distributed to end-users only in PDF format
USAGE OF DOIS FOR DATA CITATIONS

• To enable users of datasets from the archive to cite the data:
  • Every dataset for a study gets a DOI
  • Each new version of the dataset gets a new DOI
• Metadata for the study is transferred to da|ra, and from there to DataCite
• Users can find a pre-formatted citation for the dataset at each of the systems
• They can copy&paste the citation,
• But they can also read and change it according to their citation style
MACHINE ACTIONABLE PROCESSING OF METADATA

• To enable computers to process the study level metadata
  • DBK provides a webservice to access the study level metadata in DDI format
  • DataCite provides content negotiation for the metadata of the DOI record
• To access further metadata, so-called „rich metadata“, that may be kept in other files, a link to it is provided by the DataCite metadata schema:
  • relatedIdentifier of relationType „HasMetadata“
  • relatedIdentifier of relationType „IsMetadataFor“ (to link back)
• Additional information for the related identifier:
  • relatedIdentifierType
  • relatedMetadataScheme, schemeURI, schemeType
EXAMPLE IMPLEMENTATION

- DBK metadata
### EXAMPLE IMPLEMENTATION

- da|ra metadata


<table>
<thead>
<tr>
<th>DOI</th>
<th>10.4232/1.11005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>2.0.0</td>
</tr>
<tr>
<td>Resource Type</td>
<td>Dataset</td>
</tr>
<tr>
<td>Creator</td>
<td>EVS</td>
</tr>
<tr>
<td>Publication Date</td>
<td>2011-12-30</td>
</tr>
<tr>
<td>Contributor</td>
<td>Godeshi, Ilir (Center for Economic and Social Studies, Albania) (Researcher)</td>
</tr>
<tr>
<td></td>
<td>Zulehner, Paul M. (University of Vienna, Austria) (Researcher)</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Classification</td>
<td>ZA: Society, Culture, Religion and Weltanschauung</td>
</tr>
<tr>
<td>Description</td>
<td>Abstract</td>
</tr>
<tr>
<td></td>
<td>Moral, religious, societal, political, work, and family values of Europeans. Compilation of the data sets from 1981, 1990, 1999, ... see more</td>
</tr>
<tr>
<td>Temporal Coverage</td>
<td>1981-03-01 - 1981-05-31 Belgium (1st wave)</td>
</tr>
<tr>
<td></td>
<td>1982 Canada (1st wave)</td>
</tr>
<tr>
<td></td>
<td>1981-03-01 - 1981-05-31 Denmark (1st wave)</td>
</tr>
<tr>
<td></td>
<td>1981-03-01 - 1981-05-31 France see more</td>
</tr>
<tr>
<td>Geographic Coverage</td>
<td>Albania (AL)</td>
</tr>
</tbody>
</table>

...
EXAMPLE IMPLEMENTATION

- DataCite metadata
EXAMPLE IMPLEMENTATION

- DataCite content negotiation: different representations of the metadata

Other formats

- text/html
- application/x-datacite+xml
- application/vnd.datacite.datacite+xml
- application/x-datacite+text
- application/vnd.datacite.datacite+text
- application/rdf+xml
- text/turtle
- application/x-bibtex
- application/x-research-info
- application/citeproc+json
- application/vnd.citationstyle
- text/x-bibliography

```xml
<rdf:RDF>
  <rdf:Description rdf:about="http://dx.doi.org/10.4232/1.11005">
    <j:0:creator>EVS</j:0:creator>
    <j:0:publisher>GESIS Data Archive</j:0:publisher>
    <j:0:title>
    </j:0:title>
    <j:0:date>2011</j:0:date>
    <owl:sameAs>doi:10.4232/1.11005</owl:sameAs>
    <owl:sameAs>info:doi/10.4232/1.11005</owl:sameAs>
    <j:0:identifier>10.4232/1.11005</j:0:identifier>
  </rdf:Description>
</rdf:RDF>
```
EXAMPLE IMPLEMENTATION

- DBK
EXAMPLE IMPLEMENTATION

- DBKEdit: Create DOI for DDI-C document
EXAMPLE IMPLEMENTATION

- da|ra
EXAMPLE IMPLEMENTATION

- DataCite
EXAMPLE IMPLEMENTATION

• European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008)
• DataCite metadata

```
<relatedIdentifier
  relatedIdentifierType='DOI'
  relationType='HasMetadata'
  relatedMetadataScheme='DDI-C'
  schemeURI='http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd'
  schemeType='XSD'
  >10.4232/2.4804.54500</relatedIdentifier>
```

• DataCite metadata

```
<relatedIdentifier
  relatedIdentifierType='DOI'
  relationType='IsMetadataFor'
  relatedMetadataScheme='DDI-C'
  schemeURI='http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd'
  schemeType='XSD'
  >10.4232/1.11005</relatedIdentifier>
```
EXAMPLE IMPLEMENTATION

- Landing page for study: 10.4232/1.11005
EXAMPLE IMPLEMENTATION

- To programmatically access the rich metadata contained in the DDI-C file, we need direct access to the object.
- No landing page, but direct access to rich metadata: \texttt{10.4232/2.4804.54500}

\begin{verbatim}
doi:10.4232/2.4804.54500
This page represents DataCite's metadata for doi:10.4232/2.4804.54500.
For a landing page of this dataset please follow \url{http://dx.doi.org/10.4232/2.4804.54500}

Citation: GESIS - Data Archive for the Social Sciences

Resource type: Text
DDI-C Codebuch

Rights: Creative Commons Namensnennung
Creative Commons Attribution-Shared

Size
\end{verbatim}
LESSONS LEARNED

- Current implementation supports
  - Linking from study to rich metadata
  - Machine actionable processing of the metadata and rich metadata
  - Linking back from rich metadata to metadata

- Current implementation lacks
  - Machine actionable processing of the back link from rich metadata to metadata
  - Adherence to policy of using landing pages for DOI resolution
OTHER POSSIBLE SOLUTIONS

• The use of content negotiation
  • implies to have the rich metadata „inside“ the metadata (so it can be sent to the processing machine)
  • Would need explicit support from DataCite in a new metadata schema and for the rich metadata type to be returned
  • Or, a generic content type like „directAccess“ would be supported
• The use of URL instead of DOI for the link between metadata and rich metadata
  • Works only when linking items from the same provider (who has control over changing URLs)

```xml
<relatedIdentifier
  relatedIdentifierType='URL'
  relationType='HasMetadata'
  relatedMetadataScheme='DDI-C'
  schemeURI='http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd'
  schemeType='XSD'
>http://dbk.gesis.org/dbksearch/download.asp?id=54500</relatedIdentifier>

<relatedIdentifier
  relatedIdentifierType='URL'
  relationType='IsMetadataFor'
  relatedMetadataScheme='DDI-C'
  schemeURI='http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd'
  schemeType='XSD'
>https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4804&amp;db=3&amp;doi=10.4232/1.11005</relatedIdentifier>
```
THANK YOU!

Questions or comments?
Contact:

wolfgang.zenk-moeltgen@gesis.org